

## **General Disclaimer**

### **One or more of the Following Statements may affect this Document**

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

AD-A010 496

A FINITE-STATE, FINITE-MEMORY MINIMUM PRINCIPLE:  
PART II

Nils R. Sandell, Jr., et al

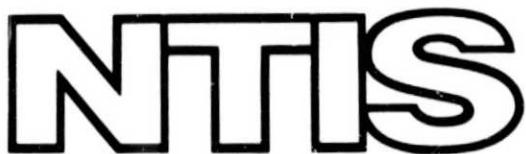
Massachusetts Institute of Technology

Prepared for:

Air Force Office of Scientific Research  
National Aeronautics and Space Administration  
National Science Foundation

1975

DISTRIBUTED BY:



National Technical Information Service  
U. S. DEPARTMENT OF COMMERCE

163110

ESL-P-592

ADA010496

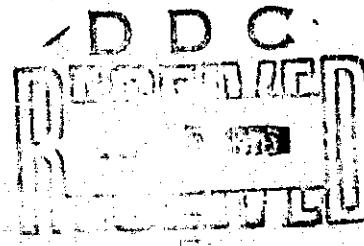
## A FINITE - STATE, FINITE - MEMORY

## MINIMUM PRINCIPLE: PART II\*

by

Nils R. Sandell Jr. \*\*

Michael Athans \*\*



Approved for public release;  
distribution unlimited.

ABSTRACT

In part I of this paper, a minimum principle was found for the finite - state, finite - memory (FSFM) stochastic control problem. In part II, conditions for the sufficiency of the minimum principle are stated in terms of the informational properties of the problem. This is accomplished by introducing the notion of a signaling strategy. Then a min-H algorithm based on the FSFM minimum principle is presented. This algorithm converges, after a finite number of steps, to a person - by - person extremal solution.

Reproduced by  
NATIONAL TECHNICAL  
INFORMATION SERVICE  
US Department of Commerce  
Springfield, VA 22151

\* This research was conducted at the Decision and Control Sciences Group of the M.I.T. Electronic Systems Laboratory with support provided under NASA Grant NGL-22-009-124, Air Force Office of Scientific Research Grant AFOSR-72-2273, and under National Science Foundation Grants GK-25781 and GK-41647.

\*\* Electronic Systems Laboratory, M.I.T. Cambridge, Mass. 02139

## UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR - TR - 75 - 0715	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD-A010 496
4. TITLE (and Subtitle) A FINITE-STATE, FINITE-MEMORY MINIMUM PRINCIPLE: PART II		5. TYPE OF REPORT & PERIOD COVERED Interim
7. AUTHOR(s) Nils R. Sandell, Jr Michael Athans		6. PERFORMING ORG. REPORT NUMBER ESL-P-592
9. PERFORMING ORGANIZATION NAME AND ADDRESS Massachusetts Institute of Technology Electronic Systems Laboratory Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 9769-01
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (NM) 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE 1975
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 28
16. DISTRIBUTION STATEMENT (of this Report)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
18. SUPPLEMENTARY NOTES  <b>PRICES SUBJECT TO CHANGE</b>		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  In part I of this paper, a minimum principle was found for the finite-state, finite-memory (FSFM) stochastic control problem. In part II, conditions for the sufficiency of the minimum principle are stated in terms of the informational properties of the problem. This is accomplished by introducing the notion of a <u>signaling strategy</u> . Then a min-H algorithm based on the FSFM minimum principle is presented. This algorithm converges, after a finite number of steps, to a <u>person - by - person extremal solution</u> .		

## I. INTRODUCTION

In this part of the paper, the development of the theory of the finite - state, finite - memory (FSFM) stochastic control problem initiated in Part I [1] is continued.

Specifically, the sufficiency of the FSFM minimum principle (which is in general only a necessary condition) is investigated. By introducing the notion of a signaling strategy as defined in the literature on games in extensive form [2], conditions under which the FSFM minimum principle is sufficient are determined. This result is interesting since it explicitly interconnects the information structure of the FSFM problem with its optimality conditions.

The paper closes with a discussion of the min-H algorithm for the FSFM problem. It is demonstrated that a version of the algorithm always converges to a particular type of local minimum termed a person - by - person extremal.

## II. SIGNALING AND SUFFICIENCY

The notion of a signaling strategy arises in the theory of Kuhn-type extensive games. According to Kuhn,<sup>1</sup> an extensive game is game tree with

- (i) a partition of the vertices with alternatives into the chance moves  $P_0$  and player moves  $P_1, \dots, P_n$
- (ii) a partition of the moves of  $P_i$  into information sets
- (iii) a probability distribution on the alternatives of the information sets of  $P_0$
- (iv) an  $n$ -tuple of real numbers for each terminal vertex.

An example of a Kuhn-type extensive game is shown in Figure 1. There is one chance move in  $P_0$  with four alternatives. Each alternative consists of the choice of an outcome of tossing two pennies. Thus each outcome occurs with probability  $\frac{1}{4}$ . There are four moves in  $P_1$ , and player one's information set is equal to  $P_1$ . Thus player one does not know the outcome of the first chance move. He has to guess if the pennies match or don't match. If he guesses correctly, he gets to keep his own penny and player two's penny (the payoff is  $(+1, -1)$ ). If he guesses incorrectly, he loses his penny to player two (the payoff is  $(-1, +1)$ ).

---

<sup>1</sup>See [3] for a complete exposition.

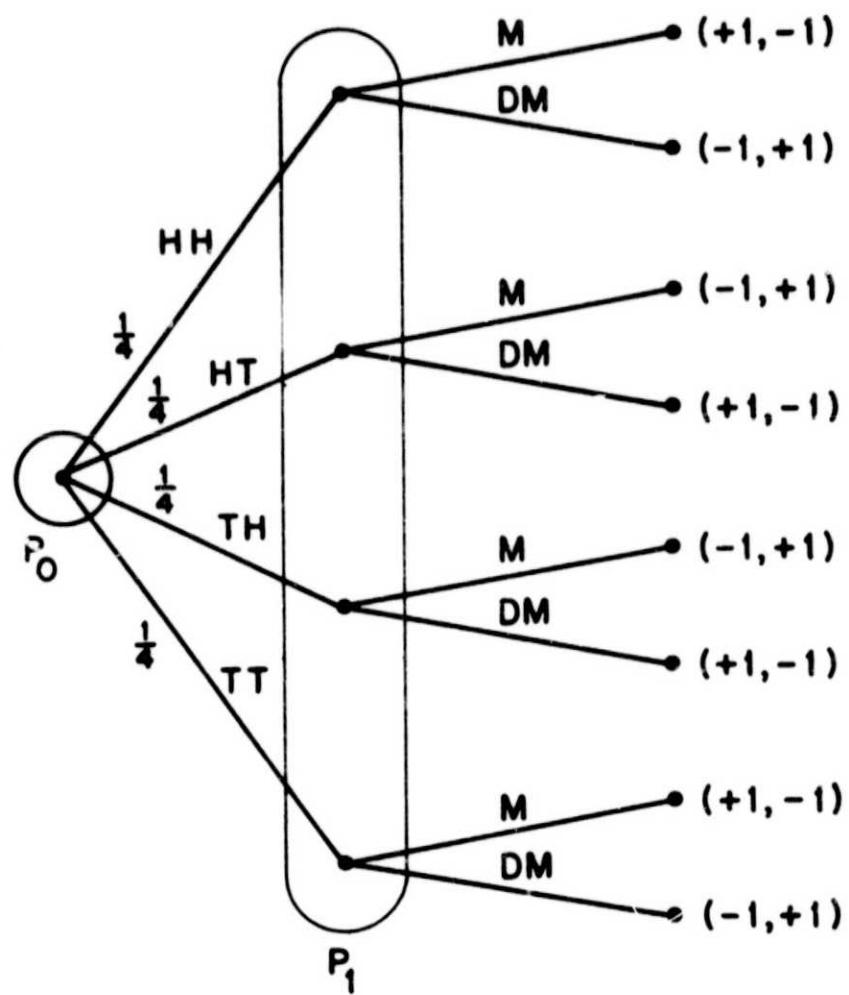


Figure 1 Matching Pennies

Every FSFM problem can be reduced to a Kuhn extensive game. It might be thought that the reduction is accomplished by identifying the player's alternatives with the controller's inputs, but this is not always possible. Suppose, for example, that  $x_0 = \{1,2\}$ ,  $u_1 = \{0,1\}$ , and  $\Gamma_1 = \{\gamma_1, \tilde{\gamma}_1\}$ , where  $\gamma_1(1) = 1$ ,  $\gamma_1(2) = 0$  and  $\tilde{\gamma}_1 = 1 - \gamma_1$ .<sup>1</sup> Clearly, the game tree for this problem must have its first seven nodes as in Figure 2, with vertices 1 and 2 in the set of moves of player one (the only player). However, it is not possible to partition  $P_1$  into information sets so that the restriction that the same alternative must be chosen for each vertex in a given information set is equivalent to the restriction that the control law must lie in  $\Gamma_1$ . The point is that restricting the control laws to lie in an arbitrary subset of  $u_t^{x_{t-1}}$  is a more general restriction than one based on information. Thus, it is in general necessary to identify the player's alternatives with the set of control laws. This is undesirable since the game does not exhibit the information properties of the FSFM problem. However, it will be shown next that the first reduction (identifying alternatives with controller inputs) is possible for FSFM problems with simple information constraint.

---

<sup>1</sup>The choice of  $\Gamma_1$  seems unnatural, but has appeared in the literature [4]. The control laws in  $\Gamma_1$  are the closed-loop control laws those in  $u_1^{x_0} - \Gamma_1$  are the open-loop control laws.

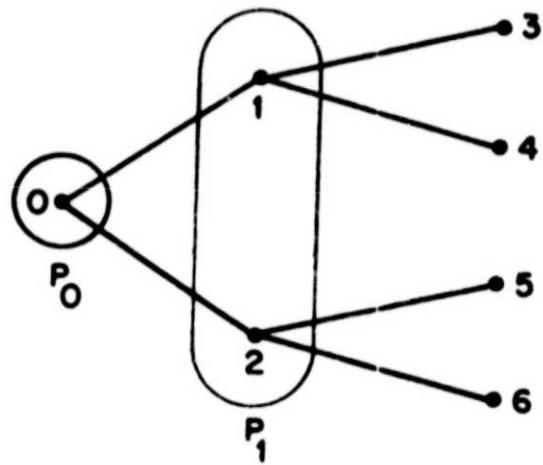


Figure 2 Game Tree for FSFM Problem

Definition 1

The FSFM problem defined by equations (1) and (2) of Part I is said to have a simple information constraint if

$$\Gamma_t = \{y_t \in u_t^{x_{t-1}} : y_t^{-1}(u_t) \subset F_{t-1}\} \quad (1)$$

for  $t = 1, 2, \dots, T$ , where  $u_t = P(u_t)$  and  $F_{t-1}$  is a subfield of  $x_{t-1} = P(x_{t-1})$ .

The reason for restricting attention to FSFM problems with simple information constraints is that these problems can be readily identified with a corresponding Kuhn model of an extensive game

Suppose that a FSFM problem with simple information constraint is given. Let the sets  $x_0, Q_1, U_1, Q_2, \dots, U_T$  have  $n_0, n_1, m_1, n_2, \dots, m_T$  elements, respectively. The rank 0 move<sup>1</sup> of the corresponding game tree has  $n_0$  alternatives. For  $1 \leq t \leq T$ , the rank  $2t-1$  move has  $n_t$  alternatives and the rank  $2t$  move has  $m_t$  alternatives. Thus every play has rank  $2T + 1$  (Figure 3).

---

<sup>1</sup>A move is a vertex of the game tree with alternatives; a play is a (terminal) vertex without alternatives. The rank of a move or play is the number of moves that precede it. See Kuhn [3] for details.

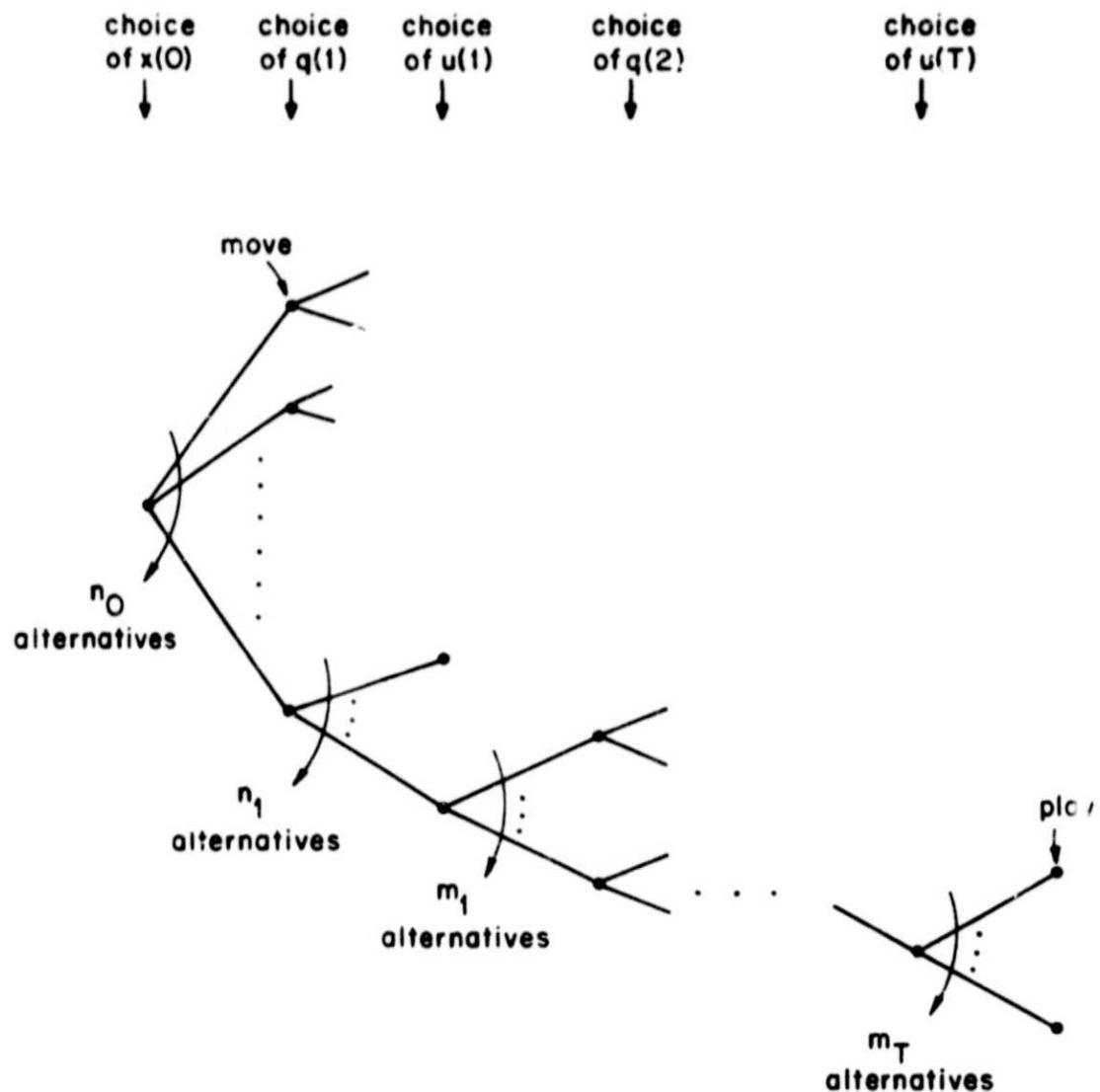


Figure 3

Game Tree for FSFM Problem With  
Simple Information Constraint

The chance moves  $P_0$  are the moves with rank 0, 1, 3, ...,  $2T-1$ , and the moves  $P_1$  of player 1 (the only player) are the moves with rank 2, 4, ...,  $2T$ . Each alternative of the initial (rank 0) move of the game tree corresponds to an element of  $X_0$ . Similarly, the alternatives of moves with rank  $2t-1$  correspond to elements of  $Q_t$ , and moves with rank  $2t$  correspond to elements of  $U_t$ .

Each information subset of  $P_0$  contains a single point of  $P_0$ . The information sets of  $P_1$  are defined by the atoms<sup>1</sup> of  $F_t$  as follows. Notice that the system equations of the FSFM problem define a map

$$s_t : X_0 \times Q_1 \times U_1 \times \dots \times Q_t \times U_t \rightarrow X_t \quad (2)$$

which takes an initial state and a sequence of inputs and gives corresponding state. Each atom  $F$  of  $F_t$  defines a set

$$\{(x(0), q(1), u(1), \dots, q(t), u(t)) : s_t(x(0), q(1), u(1), \dots, q(t), u(t)) \in F\} \subset X_0 \times Q_1 \times U_1 \times \dots \times Q_t \times U_t. \quad (3)$$

Since there is a one-to-one correspondence between the set  $X_0 \times Q_1 \times U_1 \times \dots \times Q_t \times U_t$  and the moves of order  $2t + 1$  of the game, the partition induced on  $X_0 \times Q_1 \times U_1 \times \dots \times Q_t \times U_t$  by the atoms of  $F_t$  induces a partition on the corresponding set of moves. Thus each atom  $F \in F_t$  gives rise to a single information set for player 1 containing moves of player 1. As a consequence, all the moves of given information set are of the

---

<sup>1</sup>An atom of a field  $F$  is a set  $F \in F$  such that if  $E \in F$  and  $E \subset F$ , then either  $E = \emptyset$  or  $E = F$ . The atoms of a finite field always exist and form a partition [ 5 ].

same rank. This is not surprising, since the problem is sequential [6].

To finish the specification of the game, the probabilities of the chance moves must be defined and the terminal cost specified. If an information set of  $P_0$  contains a move of rank  $2t-1$ , its alternative corresponding to  $q \in Q_t$  is chosen with probability  $p_t(q)$ . The terminal cost is determined by the fact that the plays are in one-to-one correspondence with  $X_0 \times Q_1 \times U_1 \times \dots \times Q_T \times U_T$ . Thus each play determines a complete state-control trajectory for which  $J$  can be evaluated. This value of  $J$  is the cost associated with the play.

In game theory, a strategy for player 1 is the assignment of a single alternative to each information set. For FSFM problems with simple information constraint, a control law is the assignment of a point in  $U_t$  to each atom of  $F_{t-1}$  (since  $Y_t$  is constrained to be  $F_{t-1}$  measurable). Because of the manner in which the information sets have been constructed above, there is clearly a one-to-one correspondence between the control laws of a FSFM problem with simple information constraint and **the strategies** of its corresponding extensive game form. Thus the same notation  $\gamma$  will be used to describe either a control law sequence or a strategy for the equivalent extensive game.

---

Since an equivalence has been established between FSFM models with simple information constraint and Kuhn extensive game models, the notions

of signaling strategy and perfect recall can now be precisely defined. The following definitions and propositions are stated for 1-player games, but can be easily extended to n-person games.

Definition 2 [3].

A move  $Z$  of player 1 ( $n=1$ ) is called possible when playing  $\gamma$  if it has non-zero probability of occurring when the strategy  $\gamma$  is used. An information set  $I$  for player 1 is called relevant when playing  $\gamma$  if some  $Z \in I$  is possible when playing  $\gamma$ .

Proposition 1.

A move  $Z$  for player 1 is possible when playing  $\gamma$  if and only if  $\gamma$  chooses all alternatives on the path  $W_Z$  from the origin to  $Z$  which are incident at moves of player 1.<sup>1</sup>

Proof

See reference [3], page 201.

Definition 3 [3].

A game  $G$  is said to have perfect recall if  $I$  is relevant when playing  $\gamma$  and  $Z \in I$  implies that  $Z$  is possible when playing  $\gamma$  for all  $I$  and  $\gamma$ .

Definition 4 [2].

Let  $I$  be an information set for player 1, and let  $I_u = \{\text{moves following some move in } I \text{ by alternative } u\}$ . Then  $I$  is a signaling information set

---

<sup>1</sup>All chance moves are assumed to occur with non-zero probability.

for player 1 if, for some  $u$  and some information set  $J$  of player 1,  
 $I_u \cap J \neq \emptyset$  and  $J \not\subseteq I_u$ .

Proposition 2 [2].

A game  $G$  has perfect recall if and only if player 1 has no signaling information sets.

Proof

See reference [2], page 268.

The following proposition is not valid for general games, but is a special property of 1-person (stochastic control) problems.

Proposition 3.

Let  $G$  be a 1-person game with perfect recall, and let  $I$  be an arbitrary information set of the player. If  $I$  is not relevant when playing  $\gamma$ , then the probability of any move in  $I$  is zero under  $\gamma$ . If  $I$  is relevant when playing  $\gamma$ , then the probability of any move in  $I$  is positive under  $\gamma$ . Moreover, if  $I$  is relevant under any other strategy  $\tilde{\gamma}$ , then the probabilities of any move of  $I$  under  $\gamma$  and  $\tilde{\gamma}$  are the same.

Proof

If  $I$  is not relevant when playing  $\gamma$ , then by definition no move of  $I$  is possible when playing  $\gamma$ . Thus the probability of any such move is zero when  $\gamma$  is used.

If  $I$  is relevant when playing  $\gamma$ , then every move of  $I$  is possible when playing  $\gamma$  since  $G$  has perfect recall. Thus the probability of any such move is positive when  $\gamma$  is used.

If  $z \in I$  is possible when playing  $\gamma$ , by Proposition 3.3.1  $\gamma$  must choose all alternatives on the path  $W_z$  from the origin to  $z$  which are incident at moves of player 1. All other alternatives on  $W_z$  are incident at chance moves, and the probability of  $z$  under  $\gamma$  is simply the product of the probabilities of these alternatives. But this probability is the same for  $\tilde{\gamma}$ , since  $\tilde{\gamma}$  likewise chooses all alternatives on the path  $W_z$  incident at moves of player 1.

At this point, the preceding definitions and propositions are applied to the FSFM problem.

#### Definition 5.

A FSFM stochastic control problem is said to have perfect recall if it has a simple information constraint and the corresponding extensive game has perfect recall.

#### Definition 6.

A control law  $\gamma_t$  for a FSFM problem with simple information constraint is said to be a signaling control law if an atom of  $F_{t-1}$  gives rise to a signaling information set in the corresponding extensive game.

#### Corollary 4.

A FSFM stochastic control problem with simple information constraint has perfect recall if and only if it has no signaling control laws.

#### Proof

This is a direct consequence of the definitions, the construction of the equivalent extensive game, and Proposition 2.

Theorem 5.

Suppose that a FSFM stochastic control problem with perfect recall is given. Let  $A$  be an atom of  $F_{t-1}$ . Then, for any control sequence, either the probability of all states in  $A$  is zero, or the probability of each state is a positive constant independent of  $\gamma$ .

Proof

By construction, the probability of a state  $x(t-1) \in A$  under  $\gamma$  is equal to the probability of the corresponding set of moves in the information set  $I$  generated by  $A$ . Therefore, the theorem follows immediately from Proposition 3.

The property of FSFM problems with perfect recall expressed by Theorem 5 makes it possible to strengthen the minimum principle to achieve a sufficient condition for optimality.

Definition 7.

Let the set of state probability vectors reachable at time  $t$ ,  $1 \leq t \leq T$ , when the initial state probability vector is  $\pi_0$  be denoted

$$r_t(\pi_0) = \{\pi_0 \text{ } p^{\gamma_1(1)} \text{ } p^{\gamma_2(2)} \dots \text{ } p^{\gamma_t(t)} : \gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \dots, \gamma_t \in \Gamma_t\}. \quad (4)$$

$r_t(\pi_0)$  is called the reachable set ( $r_0(\pi_0) = \{\pi_0\}$ ).

Definition 8.

Suppose that the control law sequence  $\gamma^* = (\gamma_1^*, \gamma_2^*, \dots, \gamma_T^*)$  satisfies the condition

$$\begin{aligned}
 & \pi(t-1) P^{\gamma_t^*}(t) \phi^*(t) + \pi(t-1) h^{\gamma_t^*}(t) \\
 & \leq \pi(t-1) P^{\gamma_t^*}(t) \phi^*(t) + \pi(t-1) h^{\gamma_t^*}(t)
 \end{aligned} \tag{5}$$

for all  $\gamma_t \in \Gamma_t$ , for all  $\pi(t-1) \in \pi_{t-1}(\pi_0)$  where

$$\phi^*(t-1) = P^{\gamma_t^*}(t) \phi^*(t) + h^{\gamma_t^*}(t) \tag{6}$$

for  $t = 1, 2, \dots, T$  ( $\phi^*(T) = \phi_T$ ). Then  $\gamma^*$  is said to be universally extremal.

Lemma 6.

Any universally extremal control law sequence is optimal.

Proof

The proof proceeds by induction on the number of stages  $T$ .

Suppose  $T = 1$ . Then

$$\begin{aligned}
 J(\gamma_1) &= \pi(0) h^{\gamma_1}(1) + \pi(1) \phi(1) \\
 &= \pi(0) h^{\gamma_1}(1) + \pi(0) P^{\gamma_1}(1) \phi(1)
 \end{aligned} \tag{7}$$

so that any extremal is optimal.

Suppose the lemma is valid for problems with  $T-1$  stages. It must be established that the lemma is valid for problems with  $T$  stages.

Assume that  $(\gamma_1^*, \gamma_2^*, \dots, \gamma_T^*)$  is universally extremal. It follows immediately that  $(\gamma_2^*, \gamma_3^*, \dots, \gamma_T^*)$  is universally extremal for the problem with cost

$$J(\gamma_2, \dots, \gamma_T; \pi(1)) = \sum_{t=2}^T \pi(t-1) h^{\gamma_t^*}(t) + \pi(T) \phi(T) \tag{8}$$

for any  $\pi(1) \in \mathcal{R}_1(\pi_0)$ . Therefore, by the induction hypothesis,

$$J(\gamma_2^*, \dots, \gamma_T^*; \pi(1)) \leq J(\gamma_2, \dots, \gamma_T; \pi(1)) \quad (9)$$

for all  $\pi(1) \in \mathcal{R}_1(\pi_0)$  and for all  $\gamma_2 \in \Gamma_2, \dots, \gamma_T \in \Gamma_T$ . Moreover, since

$$J(\gamma_1, \gamma_2, \dots, \gamma_T) = \pi(0) h^{\gamma_1}(1) + J(\gamma_2, \dots, \gamma_T; \pi(0) p^{\gamma_1}(1)) \quad (10)$$

it follows that

$$J(\gamma_1, \gamma_2^*, \dots, \gamma_T^*) \leq J(\gamma_1, \gamma_2, \dots, \gamma_T) \quad (11)$$

for all  $\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \dots, \gamma_T \in \Gamma_T$ .

But the assumption that  $(\gamma_1^*, \gamma_2^*, \dots, \gamma_T^*)$  is universally extremal implies that

$$\begin{aligned} J(\gamma_1^*, \gamma_2^*, \dots, \gamma_T^*) &= \pi(0) h^{\gamma_1^*}(1) + \pi(0) p^{\gamma_1^*}(1) \phi^*(1) \\ &\leq \pi(0) h^{\gamma_1}(1) + \pi(0) p^{\gamma_1}(1) \phi^*(1) = J(\gamma_1, \gamma_2^*, \dots, \gamma_T^*) \end{aligned} \quad (12)$$

for all  $\gamma_1 \in \Gamma_1$ . The lemma follows from (12) and (11).

Notice from the proof of Lemma 6 that the existence of a universally extremal control law sequence  $\gamma^*$  implies the unusual fact that the problems

$$\min_{\gamma_t \in \Gamma_t, \dots, \gamma_T \in \Gamma_T} J(\gamma_1, \dots, \gamma_{t-1}, \gamma_t, \dots, \gamma_T) \quad (13)$$

for  $\gamma_1 \in \Gamma_1, \dots, \gamma_{t-1} \in \Gamma_{t-1}$  have a common solution  $(\gamma_t^*, \dots, \gamma_T^*)$ .

Thus the existence of a universal extremal would seem to be rather unlikely.

From this viewpoint, the following property of FSFM problems with perfect recall seems rather remarkable.

Theorem 7.

Every FSFM problem with perfect recall has a universally extremal control law sequence.

Proof

The proof is constructive. The control laws  $\gamma_t$  are defined by choosing their values on the atoms of  $F_{t-1}$ .

Consider the case for  $t=T$ . Let  $A_{T-1}^i$  be an atom of  $F_{T-1}$ ,  $i = 1, 2, \dots, p$ . For simplicity of notation, suppose that  $A_{T-1}^1$  contains the first  $\ell_1$  states of  $X_{T-1}$ ,  $A_{T-1}^2$  contains states  $\ell_1 + 1$  through  $\ell_2$  of  $X_{T-1}$ , etc.

Notice that

$$\begin{aligned} \pi(T-1) P^{\gamma_T}(T) \phi(T) + \pi(T-1) h^{\gamma_T}(T) \\ = \sum_{i=1}^p \sum_{j=\ell_{i-1}+1}^{\ell_i} \pi_j(T-1) \left[ \sum_{k=1}^n P_{jk} u_i(T) \phi_k(T) + h_j u_i(T) \right] \end{aligned} \quad (14)$$

where  $n$  is the number of states in  $X_{T-1}$ ,  $\ell_0=0$ , and  $u_i(T)$  is the value of  $\gamma_T$  on the  $i$ th atom of  $F_{T-1}$ .

The decomposition (14) makes the construction of  $\gamma_T^*$  clear.

By Proposition 5, every vector  $\pi(T-1) \in r_{T-1}(\pi_0)$  either has  $\pi_j(T-1)=0$ ,  $j = \ell_i + 1, \dots, \ell_{i+1}$ , or has  $\pi_j(T-1) = \hat{\pi}_i(T-1)$ ,  $j = \ell_i + 1, \dots, \ell_{i+1}$ , where each  $\hat{\pi}_i(T-1)$  is a fixed number independent of  $\gamma_1, \dots, \gamma_{T-1}$ .

Therefore,  $\gamma_T^*$  takes the value  $u_i^*(T)$  on the  $i$ th atom of  $F_{T-1}$ , where

$$\begin{aligned}
 & \min_{u \in U_T} \sum_{j=\ell_{i-1}+1}^{\ell_i} \hat{\pi}_i^{(T-1)} \left[ \sum_{k=1}^n p_{jk}^{u(T)} \phi_k^{(T)} + h_j^{u(T)} \right] \\
 &= \sum_{j=\ell_{i-1}+1}^{\ell_i} \hat{\pi}_i^{(T-1)} \left[ \sum_{k=1}^n p_{jk}^{u_i^*(T)} \phi_k^{(T)} + h_j^{u_i^*(T)} \right]. \tag{15}
 \end{aligned}$$

The construction of the remaining  $\gamma_t^*$  is completed by applying an analogous procedure to

$$\pi(t-1) p_{\gamma_t^*(t)} \phi^*(t) + \pi(t-1) h_{\gamma_t^*(t)}. \tag{16}$$

Theorem 7 is primarily of theoretical and conceptual importance. Problems with perfect recall are more efficiently handled by deriving an equivalent deterministic problem that has a conditional probability vector for the deterministic state. (The conditioning is with respect to the field  $F_{t-1}$ .) Special cases of this procedure are implicit in the usual stochastic dynamic programming algorithm [7, 8, 9, ] and the algorithm of Sandell and Athans for the 1-step delay problem [10].

### III. A FSFM MIN-H ALGORITHM

A substantial number of numerical algorithms have been suggested for the solution of deterministic optimal control problems. The most natural of these for the FSFM problem is the min-H algorithm, which is intimately related to the minimum principle. The min-H algorithm was initially suggested by Kelley [11]. Platzman [12] has shown that the algorithm is equivalent to Howard's policy iteration method for Markovian decision processes, and has suggested its application to the imperfect state information case of that problem.

To simplify the notation, the sets  $X_t$  and  $U_t$  are assumed to have a constant cardinality for  $0 \leq t \leq T$ .

#### Algorithm (Min-H)

1. Guess  $\gamma_1^0, \gamma_2^0, \dots, \gamma_T^0$ . Set  $j = 0$ .
2. Compute  $\phi^j(T), \phi^j(T-1), \dots, \phi^j(1)$  using  $\gamma_T^j, \dots, \gamma_1^j$  in the adjoint equation  $(\phi^j(T) = \phi_T)$ . Set  $t = 1$ .
3. Choose  $\gamma_t^{j+1}$  to minimize  $\pi^{j+1}(t-1) P^{\gamma_t^j(t)} \phi^j(t) + \pi^{j+1}(t-1) h^{\gamma_t^j(t)}$ .<sup>1</sup>  
 $(\pi^{j+1}(0) = \pi_0)$
4. If  $t < T$ , compute  $\pi^{j+1}(t) = \pi^{j+1}(t-1) P^{\gamma_t^{j+1}(t)}$ .  
Set  $t = t+1$ , and go to 3.
5. If  $t = T$ , test  $J^{j+1} < J^j$ , where

$$J^j = \sum_{t=1}^T \pi^j(t-1) h^{\gamma_t^j(t)} + \pi^j(T) \phi_T$$

---

<sup>1</sup>If  $\gamma_t^{j+1}$  is not unique, choose arbitrarily but with preference for  $\gamma_t^j$  if it is in the minimizing set.

If  $J^{j+1} < J^j$ , set  $j = j+1$ ,  $t = 0$ , and go to 2.

If  $J^{j+1} = J^j$ , stop.

Theorem 8.

The preceding algorithm converges in a finite number of steps to an extremal solution.

Proof

Let  $S = \{ J(\gamma) \mid \gamma \in \Gamma \}$ . Since  $S$  is finite, its elements can be arranged in descending order,

$$S = \left( J_1, J_2, \dots, J_\ell \right), \quad J_i > J_{i+1}. \quad (17)$$

Consider the set of positive numbers

$$R = \left( J_1 - J_2, \dots, J_{\ell-1} - J_\ell \right) \quad (18)$$

and let  $\varepsilon = \inf R$ . Note that  $\varepsilon > 0$ .

Consider the difference  $J^j - J^{j+1}$  defined in the algorithm. Clearly, either  $J^j - J^{j+1} = 0$ , or  $J^j - J^{j+1} \geq \varepsilon$ . By induction, if the algorithm has not converged by step  $j$ , then

$$J^j \leq J^0 - j\varepsilon. \quad (19)$$

Therefore, eventually  $J^{j+1} = J^j$ , since  $\inf S$  is finite. But  $J^j = J^{j+1}$  implies that  $(Y_1^j, \dots, Y_T^j)$  is extremal.

Although the FSFM Min-H algorithm is guaranteed to converge in a finite number of steps, the amount of computation per step may be prohibitive, even if full advantage of the special structure of the problem is made (see [13] for a discussion and estimates of computation time). Thus modifications to the basic algorithm for special cases are of interest.

Consider the case in which

$$\Gamma_t = \Gamma_t^1 \times \Gamma_t^2 \times \dots \times \Gamma_t^k \quad (20)$$

and  $\Gamma_t^i$  consists of control laws measurable with respect to a subfield  $F_t^i$  of  $X_t$ .

Make the following notational convention:

$$Y_t : x_{t-1} \rightarrow u_t^1 \times u_t^2 \times \dots \times u_t^k \quad (21)$$

$$Y_t = (Y_t^1, Y_t^2, \dots, Y_t^k). \quad (22)$$

Then

$$\begin{aligned} & J(Y_1, Y_2, \dots, Y_T) \\ &= J(Y_1^1, \dots, Y_1^k, Y_2^1, \dots, Y_2^k, \dots, Y_T^1, \dots, Y_T^k). \end{aligned} \quad (23)$$

Definition

A sequence

$$\gamma^* = (\gamma_1^*, \dots, \gamma_T^*) = (\gamma_1^{1*}, \dots, \gamma_1^{k*}, \dots, \gamma_T^{1*}, \dots, \gamma_T^{k*})$$

is said to be a person-by-person extremal if

$$\begin{aligned} & J(\gamma_1^{1*}, \dots, \gamma_t^{i*}, \dots, \gamma_T^{k*}) \\ & \leq J(\gamma_1^{1*}, \dots, \gamma_t^i, \dots, \gamma_T^{k*}) \quad \text{for all } \gamma_t^i \in \Gamma_t^i, \end{aligned}$$

$$i = 1, \dots, k, \quad t = 1, \dots, T. \quad (24)$$

Every optimal control law sequence is a person - by - person extremal, but the converse need not be true. Clearly, the FSFM Min-H algorithm can be modified to give an algorithm that always converges to a person - by-person extremal. One possible order of minimization is

$$\gamma_1^1, \gamma_2^1, \dots, \gamma_T^1, \gamma_1^2, \gamma_2^2, \dots, \gamma_T^2, \dots, \gamma_1^k, \gamma_2^k, \dots, \gamma_T^k.$$

Thus  $k$  forward and backward sweeps of the state and costate equations are required per iteration. The number of multiplications required is considerably reduced. See [13] for details. Clearly, the person -by- person Min-H algorithm is finitely convergent to a person -by- person extremal solution.

Notice that person-by-person approach is consistent with the minimum principle approach:

1. both approaches given necessary conditions for optimality
2. both approaches are sufficient only under convexity assumptions that do not hold in general
3. An initial guess is improved, but the improvement may stop short of optimal.

These facts are consequences of the fact that the person-by-person and min H algorithms are actually both concrete realizations of orthogonal search. The Min-H algorithm minimized the cost without coordinated choice of the control laws at different times. The person - by-person Min-H algorithm minimizes the cost without coordinated choice of the control laws of the various controllers at a fixed time instant.

#### IV. SUMMARY AND CONCLUSIONS

The notion of signaling has been introduced from game theory and shown to be relevant to the FSFM problem. In fact, the signaling phenomena is of general importance in non-classical stochastic control theory. The presence of signaling makes it necessary for decentralized controllers to employ control laws with a dual purpose: simultaneous communication and control. The presence of signaling in LQG problems manifests itself in the nonlinear strategies that are optimal for these problems [1,14]. (Given the prevalence of nonlinear coding and modulation techniques in communication theory, the existence of nonlinear optimal strategies for nonclassical LQG problems is hardly surprising.) Moreover, the absence of signaling in LQG problems (in the LQG context, equivalent to the presence of Ho-Chu nesting) insures the optimality of linear strategies [15]. Thus the very special nature of the classical stochastic control problem is made clear: only the control aspect of the dual problems of communication and control need be considered.

The need to simultaneously solve a control and communication problem makes the nonclassical stochastic control problem very difficult to solve, even in the FSFM case. One approach to solution of the FSFM problem is the person - by - person Min-H algorithm sketched in Section III. Presently, evaluation of the algorithm is being carried out in the context of a highly simplified model of an ARPA-type packet switching computer communication network [16]. The primary difficulty is essentially combinatorial, since there is an explosive growth in the number of states with network size. Thus straightforward implementation of an algorithm seeking a "node-by-node" optimal routing strategy is possible only for small networks, or larger networks with an aggregated and/or merged [17] state set.

REFERENCES

- [1]. N.R. Sandell Jr. and M. Athans, "A Finite-State, Finite-Memory Minimum Principle: Part I", submitted to IEEE Trans. A.C.
- [2]. G.L. Thompson, "Signaling Strategies in n-Person Games", Contributions to the Theory of Games II, Princeton Univ. Press, Princeton, N.J., 1953.
- [3]. H.W. Kuhn, "Extensive Games and the Problem of Information", Contributions to the Theory of Games II, Princeton Univ. Press, Princeton, N.J., 1953.
- [4]. A.N. Starr and Y.C. Ho, "Further Properties of Nonzero Sum Differential Games", Journal of Optimization Theory and Applications, Vol. 3, No.4, 1971.
- [5]. P.R. Halmos, Measure Theory, Van Nostrand, N.Y. 1950.
- [6]. H.S. Witsenhausen, "On Information Structures, Feedback and Causality", SIAM J. Control, Vol. 9, 1971.
- [7]. M. Aoki, Optimization of Stochastic Systems, Academic Press, N.Y. 1967.
- [8]. K.J. Astrom, "Optimal Control of Markov Processes with Incomplete State Information", J. Math. Analysis and Applications, Vol. 10, 1965.
- [9]. C.T. Striebel, "Sufficient Statistics in the Optimum Control of Stochastic Systems", J. Math. Analysis and Applications, Vol. 12, 1965.
- [10]. N.R. Sandell Jr. and M. Athans, "Solution of Some Non-Classical LQG Stochastic Decision Problems," IEEE Trans. A.C., Vol. AC-19, No.2, 1974.
- [11]. H.J. Kelley, "Method of Gradients," in Optimization Techniques, Academic Press, N.Y.. 1962.
- [12]. L.K. Platzman, On the Iterative Design of Optimal and Suboptimal Controllers, M.S. Thesis, M.I.T., 1973.
- [13]. N.R. Sandell Jr, Control of Finite-State, Finite-Memory Stochastic Control Systems, Ph.D. Thesis, M.I.T., 1974.

- [14]. H.S. Witsenhausen, "A Counterexample in Stochastic Optimal Control," SIAM J. Control, Vol. 6, 1968.
- [15]. Y.C. Ho and K.C. Chu, "Team Decision Theory and Information Structures in Optimal Control Problems - Part I," IEEE Trans. A.C., Vol. AC-17, No. 1, 1972.
- [16]. N.R. Sandell Jr., "Data Communication Networks as Nonclassical Stochastic Control Problems: The FSFM Approach," M.I.T., Dec. 2 1974.
- [17]. R.A. Howard, Dynamic Probabilistic Systems, Vol. I, Wiley, N.Y., 1971.